Profile Based Data Mining for Agricultural Context

Lathashree.R[1]Akash.J. V[2]S. Lekhashree[3] Mohammed Rashid[4]Harshavardhan.J.R[5]
[1,2,3,4,5]Department of Computer Science &Engineering
[1,2,3,4,5]Visvesvaraya Technological University,Belagavi,Karanataka, India

Abstract**—***Data mining in agriculture is a very recent research topic nowadays. It consists of the application of data mining techniques in agriculture.Recent technologies are nowadays, able to provide a lot of information on agricultural-related activities, which can then be analyzed in order to find important information.It is also referred to as precision agriculture.In agriculture sector where farmers and agri businesses have to make innumerable decisions every day and intricate complexities involves the various factors influencing them. An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. Data mining techniques are necessary approach for accomplishing practical and effective solutions for this problem.In Data mining agriculture has been a recent topic. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions. This paper focuses on the analysis of the agriculture data and finding optimal parameters to maximize the crop production using data mining techniqu Linear Regression., MySQL. Bringing data mining technologies intoagriculturepresents a significant challenge at the same time, this technology contributes effectively in many countries economic and social development. In this work, we will study environmental data provided by precision information technologies, which represents a crucial source of data in need of being wisely managed and analyzed with appropriate methods and tools in order to extract the meaningful information.*

Keywords—Web technology, LAMP, CMS, Web Survey, Joomla

I. INTRODUCTION

In the recent years, the huge volume of real time data in the agricultural sector and its need for an efficient and effective processing, stimulate the use of novel technologies and platform to acquire, store, process, analyze and visualize large data sets for future predictions and decision making.Data Mining is an evolving term given to a wide area of data-intensive technologies in which the datasets are extremely large that dealing with them become more challenging than how it was before.Due to the critical challenges facing the agriculture sector farmers feel more forced to adopt intensive farming practices and sustainable agricultural ones, in order to increase both economic and environmental costs.

II. LITERATURE REVIEW

From farming history, how changes in asset enrichments and specialized change have actuated changes in private property rights and in the improvement of non-advertise foundations. We additionally think about the effect of advances in sociology learning and of social gifts on the supply of institutional change in paper[1]. It is used to survey veritable characteristics (cost of houses, number of calls, signify arrangements et cetera.) in perspective of perpetual variable(s). This best fit line is known as backslide line and addressed by a straight condition Y= a *X + b in paper[2]. Gross domestic product in cultivating and to measure the effect of each factor in convincing yield, in wording of versatilities. In the wake of completing a thorough review of composing, the going with factors were recognized in paper[3]. Huge information advances are identified with scattered figuring and massive information preparing flawless models. One structure that gives this point of view is Hadoop1 in paper[6].

III. PROPOSED SYSTEM

A. Linear Regression Algorithm

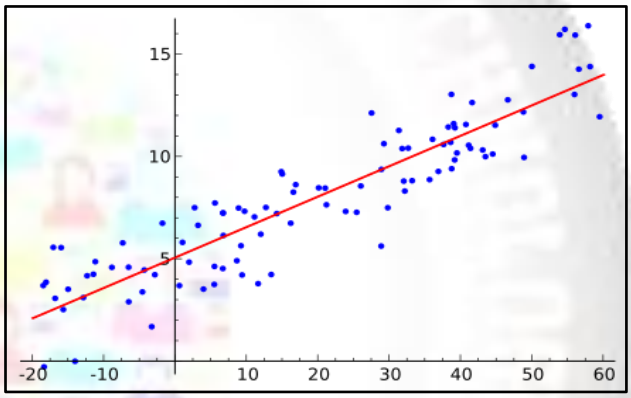

Fig. 1:linear regression algorithm

In fig1,simple linear regression is a linear regression model with a single explanatory variable.That is, it concerns twodimensional sample points with one independent variable and one dependent variable (conventionally, the x and y coordinates in a Cartesian coordinate system) and finds a linear function (a non-vertical straight line) that, as accurately as possible, predicts the dependent variable values as a function of the independent variables. The adjective simple refers to the fact that the outcome variable is related to a single predictor.

It is common to make the additional hypothesis that the ordinary least squares method should be used to minimize the residuals(vertical distances between the points of the data set and the fitted line). Under this hypothesis, the accuracy of a line through the sample points is measured by the sum of squared residuals, and the goal is to make this sum as small as possible. Other regression methods that can be used in place of ordinary least squares include least absolute deviations (minimizing the sum of absolute values of residuals) and the Theil–Sen estimator (which chooses a line whose slope is the median of the slopes determined by pairs of sample points). Deming regression (total least squares) also finds a line that fits a set of two-dimensional sample points, but (unlike ordinary least squares, least absolute deviations, and median slope regression) it is not really an instance of simple linear regression, because it does not separate the coordinates into one dependent and

one independent variable and could potentially return a vertical line as its fit.

The remainder of the article assumes an ordinary least squares regression. In this case, the slope of the fitted line is equal to the correlation between y and x corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that it passes through the center of mass (x, y) of the data points.

B. Implementation

1) Step1:

```
LetDouble[ ] year_data
double[ ]demand_data
```

2) Step2:

```
SumX= ∑ year_data
SumX2= SQRT(year_data);
SumY= ∑ demand_data
```

3) Step3:

```
doublexbar = sumx / n;
doubleybar = sumy / n;
```

Where n=no of the year.

4) Step4:

```
xxbar += (year_data [i] - xbar) * (year_data [i] - xbar);
yybar += (demand_data [i] - ybar) * (demand_data [i] - ybar);
xybar += (year_data [i] - xbar) * (demand_data [i] - ybar);
```

5) Step5:

```
double beta1 = xybar / xxbar;
double beta0 = ybar - beta1 * xbar;
predicted_value=(beta1*(present_year))+beta0;
```

A. Data Mining

Data mining is an interdisciplinary subfield of computer science. It is the computatinalprocess of discovering patterns in large data sets involving methods at the intersection ofartificial intelligence,machine learning, statistics, and database systems.The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.Data mining means collecting relevant information from unstructured data. So it is able to help achieve specific objectives. The purpose of a data mining effort is normally either to create a descriptive model or a predictive model .A descriptive model presents, in concise form, the main characteristics of the data set. The purpose of a predictive model is to allow the data miner to predict an unknown (often future) value of a specific variable; the target variable.The goal of predictive and descriptive model can be achieved using a variety of data mining techniques as shown in fig 2

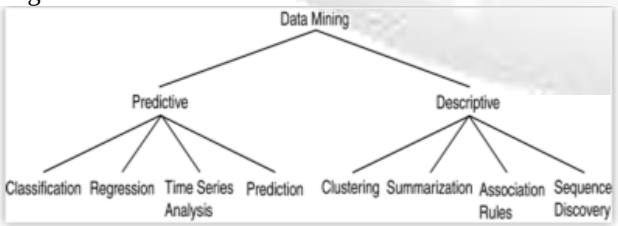

Fig. 2:Data Mining Models

B. DATA SET

The Dataset incorporates the vegetable demand,crop request of the earlier years. It likewise gives the rancher about the vegetable yield, dataset month to month savvy and furthermore about the locale astute crops,season shrewd cropswhich incorporates seasons like karif, summer, winter. The dataset are alluded from the site http://eands.dacnet.nic.in/latest_20011.htm which gives the dataset of the yields, harvests of the earlier years information. Where the agriculturist will have the capacity to check the request of his harvests and develop his yields in light of it.And properties includethe past 10 year dataset like

- Vegetable demand
- Crops demand
- Vegetable yield
- Dataset monthly wise
- Vegetable demand dataset monthly wise
- District wise crops
- Data season wise(Karif,Summer,Winter)

IV. CONCLUSION

Main Conclusion of the system is giving the suggestion of the farmers using Regression Technique. In this work, we will study environmental data provided by precision agriculture information technologies, which represents a crucial source of data in need of being wisely managed and analyzed with appropriate methods and tools in order to extract the meaningful information. So that the data obtained will be useful for the farmers to select the required crop based on the demand in the market and can get good yield and also can except profit for the grown crops. And also it will increase the economy of the country and farmers also so this tool will be very helpful for all human kind and also will effect the growth and stability of the country and all the crops. And also this will be reduce the time burden of the farmers and will easily get the suggestion so that the farmers of our country will not be in any confusion and will get the required profit and also it will make other farmers to take initiative to grow more and better crop in the future. So this tool will be very useful for the farmers and also it takes very less time to predict and give the result and also the SMS will be sent directly to registered mobile number so that the farmer need not have internet also in the phone so that he gets normal message and can take actions. So this tool can be the future app for the farmers for getting better and efficient outcome of the crop.

REFERENCES

[1] Y. Hayami and V. W. Ruttan, Agricultural development: aninternational perspective, 1st ed. Baltimore, London: The JohnsHopkins Press, 1971.

[2] S. Ray, “Essentials of Machine Learning Algorithms (with Pythonand R Codes),”in Analytics Vidhya, 2015. [Online].Available:https://www.analyticsvidhya.com/blog/2015/08/commonmachine-learning-algorithms/. Accessed: Nov. 29, 2016.

[3] P. Vinciya, Dr. A. Valarmathi, "Agriculture Analysis for NextGeneration High Tech Farming in Data Mining," International Journalof Advanced Research in Computer Science and SoftwareEngineering(ijarcsse), vol. 6, issue. 5, pp. 481-488, May.2016.

[4] Mohanraj I and J.Naren “An architectural Framework for E-AgriculturalSystem” International Conference on

Computing for Sustainable GlobalDevelopment(INDIACom),2016.

[5] N. KUMAR, G. P. O. REDDY, S. CHATTERJI, and D. SARKAR,"An application of ID3 Decision Tree Algorithm in land capabilityclassification," Nagpur 440033, India, 2012, pp. 35–42.

[6] Rosangela de Fatima Pereira, Marcelo Risse de Andrade, ArturCarvalhozucchi,karenLangona, Walter Akio Goya Nelson Mimura Gonzalez,Tereza CristinaMelobrito de Carvalho, Jan-Erimangs and Azimehsefidcon "Distributed processingfrom large scale sensor network using Hadoop" ,IEEE Interenationalcongress,2013.

[7] Mukesh Kumar and Prof.Mayuranagar "Big Data analytics in agriculture and distribution channel" Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication(ICCMC)

[8] Purnima Shah, Deepak Hiremath and Sanjay Chaudhary "Big Data Analytics Architecture for Agro Advisory System "IEEE 23rd International Conference on High Performance Computing Workshops,2016

[9] Olakunle Elijah, IgbafeOrikumhi , Tharek Abdul Rahman ,Suleiman AliyuBabale and Stella IfeomaOrakwue "Enabling Smart Agriculture in Nigeria: Application of IoT and Data Analytics" IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON),2017

[10] Rosangela de Fátima Pereira*, Marcelo Risse de Andrade*, ArturCarvalhoZucchi*, Karen Langona*, Walter Akio Goya* Nelson Mimura Gonzalez*, Tereza Cristina MeloBrito de Carvalho*, Jan-EriMångs† and AzimehSefidcon "Distributed processing from large scale sensor network using Hadoop" ,IEEEInterenational congress, ,2013

[11] Mohanraj I and J.Naren "An Architectural Framework for E-Agricultural System" InternationalConference on Computing for Sustainable Global Development(INDIACom),2016

[12] YogeshGandge and Sandhya "A Study on various Data Mining Techniques For Crop Yield Prediction" International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques(ICEECCOT),2017

[13] YashSanghvi, Harsh Gupta, HarmishDoshi, DivyaKoli, AmoghAnsh ,Umang Gupta" Comparison of Self Organizing Maps and Sammon's Mapping on agricultural datasets for precision agriculture" IEEE Sponsored 2nd International Conference on Innovations in Information ,Embedded and Communication systems (ICIIECS)2015